

# Fake News Detection on Social Media Using Regional Convolutional Neural Network Algorithm

P. Swetha<sup>1\*</sup> & Mrs. E. Priyanka<sup>2</sup>

<sup>1</sup>UG Student, <sup>2</sup>Assistant Professor, <sup>1,2</sup>Department of CSE, IFET College of Engineering, Villupuram, India.  
Corresponding Author Email: swethapanneer55@gmail.com\*



DOI: <https://doi.org/10.46431/MEJAST.2023.6218>

Copyright © 2023 P. Swetha & Mrs. E. Priyanka. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Article Received: 27 April 2023

Article Accepted: 19 June 2023

Article Published: 29 June 2023

## ABSTRACT

As social media and content-sharing platforms have evolved; misinformation and fake news have spread like wildfire, leading people to believe harmful misinformation. In this way, they can influence public opinion, spread fear, and drive people insane. Fake news identification is a current area of research aimed at determining whether content is genuine. In addition, this has significantly increased the daily amount of information on the Internet. Information can go viral in a matter of seconds thanks to social media, which has changed the way we share and process news. Everyone now relies on many online news sources because the internet is so widely used. News quickly disseminated across millions of users in a very short period of time along with the increase in the use of social media platforms like Facebook, Twitter, etc. The spread of fake news has far-reaching effects, including the formation of skewed beliefs and the manipulation of election results in favour of particular politicians. Moreover, spammers utilise alluring news headlines as click-bait for their adverts in order to make money. To provide more accurate predictions, RCNN models are trained to identify language-driven features according to content properties. This model addresses this using an efficient feature selection method.

**Keywords:** Machine learning; RCNN; Fake news; Social media; Language-driven features.

## 1. Introduction

Social networks are becoming more and more common in our professional and personal lives. More and more people are using social media instead of traditional media to find and consume news. Important news is often first reported on social media before being shared on traditional media such as television and radio. Due to the spread of news on social networks, users rarely check the accuracy of the material they post on social networks. Hoaxes, rumours, urban legends, and fake news are examples of inaccurate and twisted information commonly found on social media. Also, it is difficult to stop the spread of misinformation that is already widely circulated. This widespread use of can raise suspicion and affect people's ability to distinguish between true and false news. Numerous methods for detecting fake news are known in the literature.

Fake news is becoming increasingly difficult to identify because people with bad intentions write it so convincingly that it is difficult to distinguish it from the real news. We use a crude method of looking at news headlines and trying to determine if they are fake. Today's fake news ranges from satirical articles to government, from news that is fabricated and intended as propaganda, it causes a variety of problems. Our society suffers from a growing problem of fake news and a lack of trust in the media. One of his will fully deceptive articles is “phoney news”. But recently, social media agitation has changed that definition. Following the media attention, Facebook has been the target of intense criticism. They already allow users to report fake news on their site. In addition, they say they are working on a tool that will automatically detect fake articles. It's a challenging task. Fake news can be found on both ends of the political spectrum, so the algorithm must be politically fair. Ideally, equal consideration should be given to both legitimate news sources on both ends of the spectrum.

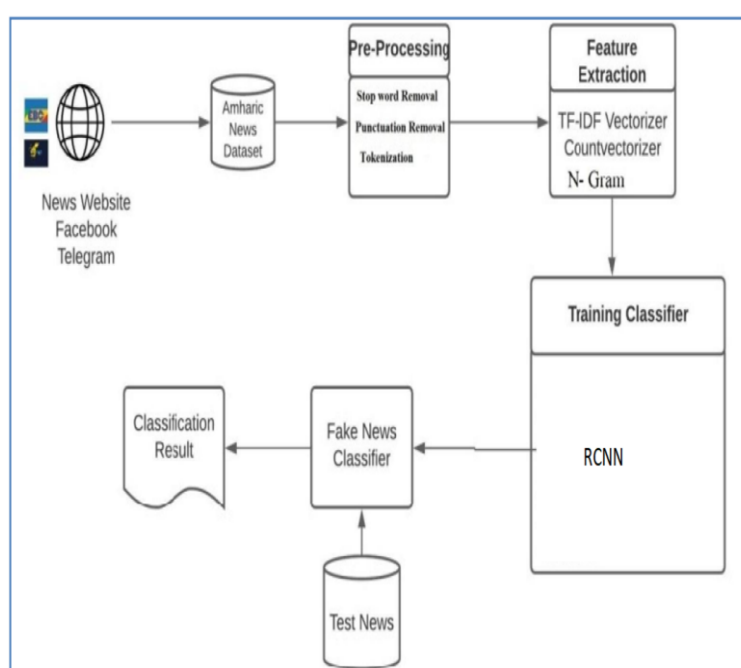
In addition, the issue of legitimacy is also a challenge. But to solve the problem, it's essential to first understand what fake news is. The advent of social media, especially Facebook news feeds, has increased the frequency of fake news that was once printed everywhere.

## 2. Related Works

Forecasting information credibility in time-sensitive social media [1]. Call attention to rumours: Deep attention-based recurrent neural networks for rumour detection in the early stages [2]. Convolutional neural networks for stance identification and rumour verification at several-2017 task 8 is something Ikm is working on [3]. USFD at semeval-2016 task 6: Any-target stance detection on Twitter with autoencoders [4]-[6]. Fake News Detection using Bi-directional LSTM-Recurrent Neural Network [7]. EANN: Event Adversarial Neural Networks for Multi-Modal [8]. Fake News Detection on Social Media: A Data Mining Perspective [9]. CSI: A Hybrid Deep Model for Fake News Detection Identifying the signs of fraudulent accounts using data mining techniques [10]. Automatic Deception Detection: Methods for Finding Fake News [11]. An Introduction to Bag-of-Words in NLP [12].

## 3. Proposed Work

Owing to the intricacy of finding fake news on social media, it is clear that a workable solution must include numerous components to tackle the problem head-on. Because of this, the suggested system combines semantic analysis. The suggested system is made up exclusively of machine learning techniques. The three-part system combines Machine Learning algorithms into plain language segments. Notwithstanding these limits, machine literacy has been crucial in the information's bracketing. This design use RCNN techniques to identify false and made-up news. By requiring deep literacy, the limitations of similar strategies and extemporisation are also examined. Successful identification of bogus news and posts utilising vibrant in projected arrangement, provincial convolutional interconnected system model is used to discover the fake information. It is more correct than the composite semantic model. Provincial convolutional interconnected system treasures everything under the footing of framework or beliefs. It finds either the crop concerning this method is fake a suggestion of correction and more the type of the revelation. It more finds the fake revelation apart from the prepared basic document file by manipulative the rate of reliable score.



**Figure 1.** Architecture of proposed system



place work. Online revelation maybe composed from various beginnings, to a degree press service homepages, search tools, and public news websites.

However, manually deciding the truth of information is a disputing task, ordinarily needing annotators accompanying rule knowledge the one acts cautious study of claims and supplementary evidence, circumstances, and reports from authorized beginnings. Generally, information dossier accompanying annotations maybe assembled in the following habits: Expert correspondents, Fact-hindering websites, Industry detectors, and Crowd culled traders. However, skilled are no coordinated standard datasets for the fake information discovery question. Data assembled must be pre-processed- that is to say, uncluttered, reconstructed and joined before it can bear preparation process. The Amharic text classification dataset that consists of more than 50k news articles that were categorized into 6 classes. This dataset is made available with easy baseline performances to encourage studies and better performance experiments.

The most well-liked words (or tags) discovered in free-form text are displayed visually in a tag cloud, sometimes referred to as a word cloud, wordle, or weighted list. Collocations and tags should be proportionately larger to how frequently they appear in your content. Its goal is to make it easier for the user to navigate the website and find what they're looking for. It is a novel approach to enhancing a website's usability and navigability while enabling us to computer the content through tags.

## 4. Experimental Results

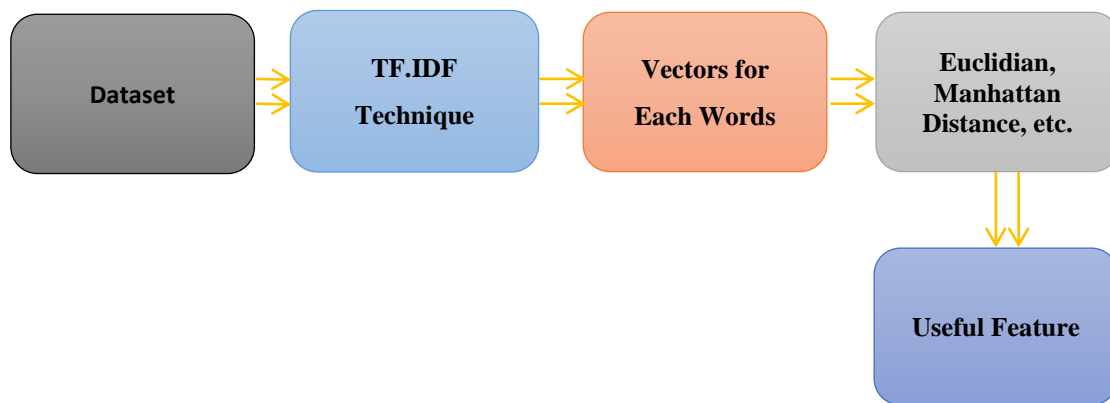
### (A) Extraction of Features

Although structural features like strings and graphs can also be utilised in machine learning, numerical features are typically the most common. In the context of our work, characteristics reflect several properties of the news story, such as its title, the amount of words, feeling, etc. Fact Check (FC), Reputation (Rep), and Coverage make up our own suggested set of features for fact-verification (CV). If you look for tf-idf right now, you could be familiar with feature extraction. The ability to convey the significance of a given word or phrase in a given document is one of the most crucial approaches utilised for information retrieval.

Let's consider a string or bag of words (BOW) as an example. If we need to extract information from it, we can employ this strategy. The frequency of the word in the corpus is frequently offset by the frequency of the word in the document, which helps to account for the fact that some words appear more frequently than others overall.

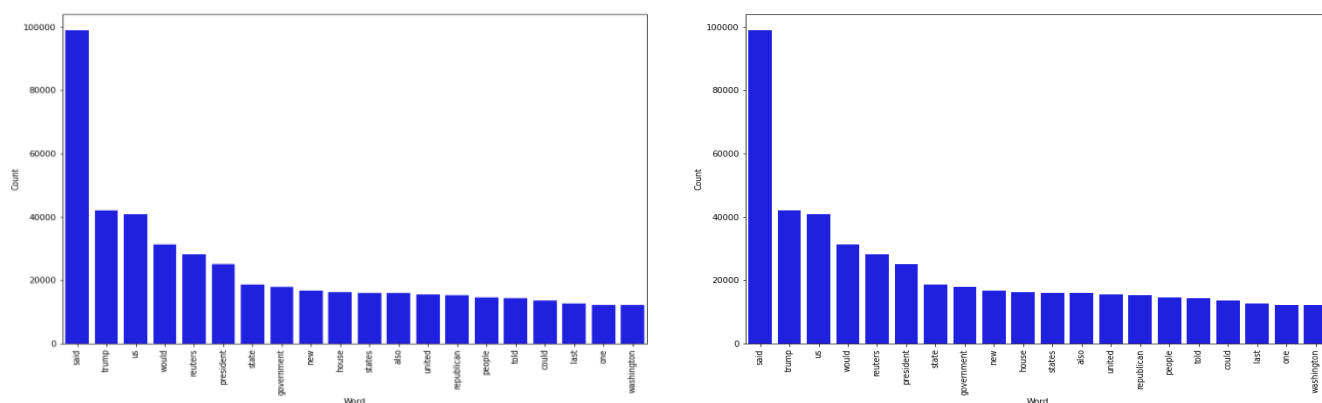
The tf-idf value rises in proportion to the number of times a word appears in the text. Two statistical techniques are used by TF-IDF, the first of which is term one is called phrase Frequency, while the other is called Inverse Document Frequency. The phrase “phrase frequency” refers the proportion of the overall occurrences of a particular term (t) in a text (doc) to the total occurrences of of words in the document.

The amount of information a word delivers is gauged by its inverse document frequency. It gauges how important a specific word is across the whole text. IDF display the word's frequency across all documents. The formula to calculate TF-IDF is  $tf * idf$ .  $Tf*Idf$  does not immediately transform raw data into meaningful characteristics. Initially, it turns raw strings or datasets into vectors, with each word having its own vector. Then, we'll employ a specific method, such as Cosine.

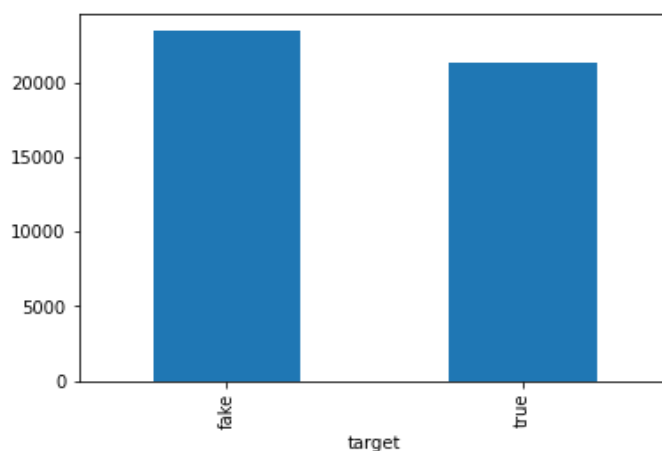


**Figure 3.** Steps for the process

The Extract N-Gram Features from Text component, and you should attach the dataset containing the text you want to process. Choose a string-type column that contains the text you wish to extract using the Text column option. You are only able to process one column at a time due to the verbose results. For the purpose of establishing a new list of n-gram features, set the vocabulary mode to Create. Indicate the largest n-gram to be extracted and stored by setting N-Grams size. Three will generate unigrams, bigrams, and trigrams, for instance. The document feature vector construction and vocabulary extraction processes are described by the weighting function.



**Figure 4.** Graph representation of word count



**Figure 5.** Output of fake and true news

Here we applied Regional Conventional Neural Network to find fake news. Our system depended heavily on this functional requirement. The system must get a genuine news article URL from the user, from which it pulls content,

in order for all of its parts to function perfectly. The web crawler will produce an exception if it does not receive a news article URL from the system. In order to meet this need, we utilised a form input of the URL type, which only accepts a URL as input. Also, we used exception handling to catch the exception in the event that the provided URL does not lead to a news article. Our project's hardest problem was this one. We simply required the pertinent content from the page source, on which our algorithm applied Natural Language Processing to create feature vectors, to categorise the news story as bogus or credible. Making a universal scraper that works for all news websites was extremely challenging in this case.

## 5. Conclusions

Some people even claim that Donald Trump became president as a result of some bogus twitter, which confuses people about whether to believe or not. We are developing a method for machine learning to address this issue. Our effective scraper takes the headline and body text from the news article, and using Natural Language Processing (NLP), we retrieved 38 features and applied the regional convolutional neural network method to determine if the news is true or false. With everyone having easy access to social media sites like Facebook and Twitter, this online application provides a solution to a critical issue. Our web application offers users an easy way to easily access news from social media, which has a significant impact on how people think. to evaluate the reliability of any news report. Our programme can be quite helpful in the real world, as seen by the accuracy of 98%. The system also includes a user feedback mechanism that allows a user to vote if the news was accurately predicted, even though there is a probability that our web application would forecast the news incorrectly. After a month or two, user votes will be carefully reviewed, and if the forecast was inaccurate, the outcome will be manually modified. The efficiency and accuracy of the application can be improved by using these projected news articles to train the machine learning models. Through time and user input, we can enhance the accuracy and usability of our programme.

## **Declarations**

### **Source of Funding**

This study did not receive any grant from funding agencies in the public or not-for-profit sectors.

### **Competing Interests Statement**

The authors have declared no competing interests.

### **Consent for Publication**

The authors declare that they consented to the publication of this study.

### **Authors' Contribution**

Both the authors took part in literature review, research, and manuscript writing equally.

## **References**

- [1] C. Castillo, M. Mendoza, et al. (2013). Forecasting information credibility in time-sensitive social media. *Internet Research*, 23(5): 560–588.



- [2] T. Chen, L. Wu, X. Li, J. Zhang, H. Yin, & Y. Wang (2017). Call attention to rumours: Deep attention-based recurrent neural networks for rumour detection in the early stages. Preprint ArXiv:1704.05973.
- [3] Z.-Y. Liu, Y.-C. Chen, and H.-Y. Kao (2017). Convolutional neural networks for stance identification and rumour verification at semeval-2017 task 8 is something Ikm is working on. Sem Eval Proceedings, ACL.
- [4] I. Augenstein, A. Vlachos, & K. Bontcheva (2016). Usfd at semeval-2016 task 6: Any-target stance detection on Twitter with autoencoders. In Sem Eval@NAACL-HLT, Pages 389–393.
- [5] S.B. Doug Sibley & Yuxia Pan Talos (2016). Available online at <http://blog.talosintelligence.com>.
- [6] B.S. Andreas Hanselowski, Avinesh PVS., & F. Caspelherr (2017). Team athene on the fake news challenge.
- [7] Bahad, P., Saxena, P., & Kamal, R. (2019). Fake News Detection using Bi-directional LSTM-Recurrent Neural Network. Procedia Computer Science, 165: 74–82.
- [8] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., & Gao, J. (2018). EANN. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- [9] Kai Shu, et al. (n.d). Fake News Detection on Social Media: A Data Mining Perspective.
- [10] Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI: A Hybrid Deep Model for Fake News Detection. arXiv. doi: <https://doi.org/10.48550/ARXIV.1703.06959>.
- [11] Conroy, N.K., Rubin, V.L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. In Proceedings of the Association for Information Science and Technology, 52(1): 1–4.
- [12] J. D'Souza (2018). An Introduction to Bag-of-Words in NLP. Available: <https://medium.com/greyatom/an-introduction-to-bag-of-words-in-nlp-ac967d43b428>.
- [13] G. Bonaccorso (2017). Artificial Intelligence, Machine Learning, and Data Science. Available: <https://www.bonaccorso.eu/2017/10/06/mlalgorithms-addendum-passive-aggressivealgorithm>.